Discourse Analysis and Document Analysis: SGML, Scholarly Publishing and Structuralist Theory

D. Grant Campbell
School of Library and Information Studies
Dalhousie University
grant.campbell@dal.ca

This paper studies the SGML Document-Type Definition for journal articles, created by Elsevier Science Publishers. Using the structuralist theory of Roland Barthes to analyze the various levels on which discourse is produced, the paper concludes that DTDs define many features of discursive structure by determining the hierarchy and grammar of elements and element sequences, and by distinguishing between mandatory and optional elements. However, features of the Elsevier DTD suggest that a DTD also needs built-in flexibility, which enables the author to define a discursive structure within an individual article; furthermore, successful scholarly publication depends also on correctly rendering the semantic units at the DTD stage, and then formatting them effectively to communicate the necessary narrative codes.

SGML and Scholarly Publishing

Standard Generalized Markup Language was declared the ISO Standard for document interchange in 1986. It is a document description language, rather than a processing language; instead of providing formatting instructions, SGML seeks to describe the structure of a document in a way which will facilitate the easy interchange of the document between multiple systems. By creating a distinction between structure and format, and between purpose and appearance, SGML enables us to identify the logical units of a document, apart from any specific instructions concerning the way those units will appear on screen or on the page. As defined by Eric van Herwijnen (1994, 23), "SGML is not a 'markup language' in the sense of troff, TeX or LaTeX: it contains the rules for creating an unlimited variety of markup languages, but does not concernitself with the formatting of marked-up documents." In this sense, SGML facilitates the

creation and definition of any number of document structures. Yuri Rabinsky, in his Forward to Charles Goldfarb's *SGML Handbook* (1990, xi), suggests that this ability to accomodate almost infinite variety is the source of SGML's success to date: "everything about its system . . . implies that a developer has the tools to build exactly what is required to indicate the internal structure of any type of information."

SGML is emerging as a powerful tool for scholarly publishing in electronic form, for two primary reasons. First, the use of non-proprietary code makes SGML the ideal language for using documents in more than one format, and on more than one system. A single document instance can be styled and mounted on any number of systems, making it possible to publish that document simultaneously in electronic and in paper form. Second, SGML enables publishers to create specific markup to facilitate precise searching: "the fact that a structured representation of the document is being maintained allows more context sensitive retrieval operations to be defined" (Macleod 1990, 201). Projects such as the FIS Electronic Journals Project show how electronic journals can be marked up in SGML to facilitate searching, and then processed through PERL scripts for mounting on the Web.[1] Indeed, many SGML tags in document headers can be mapped to MARC fields, which, with the emergence of metadata standard, can facilitate the quick creation of basic bibliographic records.[2] This combination of portability and searchability has made SGML an attractive proposition to scholarly publishers, and Elsevier Science Publishers has made considerable strides in developing an SGML publishing program for its scientific journals.

In addition to its distinctions between structure and appearance, SGML also operates on another fundamental distinction: between the document *type* and the document *instance*. Thedocument type is defined in the Document Type Definition, or DTD: this document, written in SGML, defines the structure of a particular *kind* of document: an article, for instance, or a memo, a bibliography, or an abstract. The document instance contains the data which constitutes a particular document: the words of the article, memo, bibliography or abstract, along with the markup tags which were defined in the DTD. As Rubinsky makes clear (1990, x), the "document" consists of *both* the declarations and the marked up content, the definition and the instance. And since SGML was conceived primarily as a means of document interchange, the question has arisen: to what extent

should we be using common DTDs? A common DTD is not a strict necessity for document interchange; as long as a document instance is accompanied, in interchange, by its DTD, the document can be effectively transferred and processed. Poppelier (1992, 3), at Elsevier Science Publishers, however, has argued that scientific publishing would benefit considerably from the use of a common DTD, which would enable authors to use one markup system, regardless of where they sent their paper, and publishers to store bibliographic information in separate bibliographic databases. The Elsevier Science DTD for a full-length article, which has evolved from its beginnings in 1993 to version 4.1.0 in 1997, constitutes an attempt to provide just such a common DTD. As such, the Elsevier Science DTD provides a template for scholarly electronic publishing: it establishes an essential structure of a scholarly article, and thereby establishes the rules and procedures by which this specific form of discourse can take place.

The emergence of a standard DTD for scholarly publishing, however, raises certain questions. Since the DTD defines a document's structure, how can we understand and evaluate this structure of a scholarly article? Furthermore, to what degree does this defined structure define and limit what can be said within it? This paper does not attempt to answer thosequestions definitively; rather, it seeks to find new ways to approach these questions, by drawing on structuralist theory, as it developed in the fields of literary studies and critical theory. In particular, it uses a seminal text of structuralist theory, Roland Barthes's 1966 article, "Introduction to the Structural Analysis of Narratives," to expose some of the features of structure in textual discourse, and to suggest some of the directions in which an evaluation of DTDs might proceed.

Structuralism

The choice of Barthes and "The Structural Analysis of Narratives" as a means of analyzing SGML DTDs requires some justification. To begin with, Barthes is hardly a new face on the literary studies scene, and his approaches have been eclipsed in recent years by poststructuralist and new historicist approaches . Furthermore, Barthes is discussing narrative, as a specific discourse; his findings, therefore, will translate only uneasily to the discourse of scholarly communication. Finally, the structuralist approach which Barthes practised

in this article was a passing phase: "Barthes was eager to promote his French brand of STRUCTURALISM for only a few years before he rejected most of its methodological assumptions" (Baraté 1994, 68). Structuralism, as an approach to the study of literary texts, was rapidly superseded in the seventies and eighties, not only by the rise of post-structuralism, but also by the trends of discourse theory.

In particular, Barthes's approach in this essay is rooted in one of the fundamental concepts of structuralism: the linguistic distinction, first drawn by Ferdinand de Saussure, between *langue* and *parole*: between the broad range of utterances that a language makes possible, and the individual speech acts that are uttered using that language. Barthes, in this essay, concentrates on*langue*: the range of possibilities that are possible in narrative, rather than in the study of concrete narrative utterances. In so doing, he stands well outside the current practices of discourse analysis. As developed on the foundations of Bakhtin and Foucault, discourse analysis takes issue with the *langue/parole* distinction, arguing that discourse should be studied within the social, political context of its utterances: "a theory of discourse . . . implies a theory of society, most particularly a theory of power, legitimacy, and authority" (Dillon 1994, 211). Clare Beghtol, discussing the distinction in text linguistics between virtual and actual language systems, suggests that text linguistics remains focussed on actual, rather than virtual systems: "text linguists prefer to study the actual linguistic behaviour that has purposefully occurred as a result of the wish of aspeaker or writer to communicate" (1986, 88). Within such a context, Barthes's efforts to discover a fundamental *langue* of narrative might seem outmoded or misguided.

It is this very emphasis on the "unfashionable" part of the *langue/parole* distinction which makes Barthes useful for the study of SGML, particularly the study of the creation of document-type definitions. Standard Generalized Markup Language rests on a fundamental distinction between the Document Type Definition and the Document Instance, and this distinction roughly corresponds to the *langue/parole* distinction posited by Saussure and linguistic theory. The DTD defines the essential structure of a document type, thereby establishing the rules and structural possibilities of utterance; the document instance constitutes an individual text, or utterance, which is marked up in conformance with

that specific range of structural possibilities and markup codes. A study of DTDs, then, is a study of the conscious and methodical creation of a *langue*: a set of rules and structures within which any number of utterances can theoretically take place. Such a study would benefit, therefore, from a theory of *langue*: the rules and codes whichgovern the creation of discourse.

Narrative, Discourse and Structure

Barthes's theory rests upon a supposed homological relationship between linguistics and narrative: the principles governing the creation of sentences, he argues, are essentially the same as those governing the creation of discourses. This similarity is obscured, ironically, by the fact that natural language is used as the vehicle for the creation of the broader system of meaning encompassed by narrative. This similarity should pose no problems for us in information studies; we are accustomed to studying the use of natural language to create artificial languages, in the form of controlled vocabularies and thesauri, for subject access. However, SGML complicates the tension between language and discourse by introducing a third factor. Yuri Rubinsky describes SGML as a metalanguage: a language which governs the way in which texts can be created (1990, xi). We have, then, a triple tension between language, metalanguage and discourse. In the case of the Elsevier article, the scholarly discourse is created through a combination of natural language (the words of the author) and a metalanguage: the markup tags defined in the DTD.

Barthes argues that the elements of narrative exist on different levels, and that meaning is produced through the interaction between the elements:

> A sentence can be described, linguistically, on several levels (phonetic, phonological, grammatical, contextual) and these levels are in a hierarchical relationship with one another, for, while all have their own units and correlations . . . no level on its own can produce meaning. A unit belonging to a particular level only takes on meaning if it can be integrated in a higher level. (1966, 257-58)

By envisioning narrative as a collection of hierarchical levels, Barthes introduces a tension that will be familiar to students and scholars of classification: a tension between distribution andintegration, between relationships that occur among units within a level of a hierarchy, and those among units on different levels. Distributional relationships follow a syntagmatic pattern, which is conditioned by sequence, and by the concept of "metonymy," while integrational relationships follow a paradigmatic pattern, which is conditioned by "consequence" rather than "consecution," and by the concept of "metaphor." For Barthes, understanding a narrative entails some grasp of its overall synchronic structure, in addition to an experience of its sequential unfolding: "To understand a narrative," he claims, "is not merely to follow the unfolding of the story, it is also to recognize its construction in 'stories,' to project the horizontal concatenations of the narrative "thread" onto an implicitly vertical axis; to read (to listen to) a narrative is not merely to move from one word to the next, it is also to move from one level to the next" (259).

As Ian Macleod has pointed out (1990, 198), SGML views a document as a hiearchical structure, in contrast to the relational model of database design. Like all DTDs, therefore, the DTD for an Elsevier article provides a hierarchical structure, in which the root element, "art" is defined as a sequence of sub-elements, which are in turn defined as series of sub-sub-elements:

Table 1: A Partial Diagram of the Elsevier DTD Article Hierarchy
© Elsevier Science Publishers, 1993-1997

Root Element	ART
Sub-Elements	COPYRIGHT, DOCHEAD, DOCTOPIC, FM, BDY, BM
Sub-Elements of the Element, "FM"	ALT, PRS, DED, AUG, RE, RV, ACC, MISC, ABS

The discursive significance of each element is determined partly by its relations with other elements in the parent element, and partly by its integration into a broader element, just as narrative, for Barthes, consists of sequences of units, which in turn become units in broadersequences. The element "RV," for revision date, is part of a sequence depicting the front matter of an article, which in turn

forms a sequence known as "article." It derives its significance for the reader partly by being collected with a received date and an accepted date, and also by its presence in a scholarly article.

However, SGML complicates matters, because structure is by no means a single, uniform phenomenon in a text, and documents can be used for many different purposes. Literary scholarship, for instance, can find multiple structures within a single text: a novel could be meaningfully defined in terms of its chapters, its characters, or by the page divisions of the first edition, depending on the use to which it is put. As the Text Encoding Initiative has discovered,

> These textual structures overlap with each other in complex and unpredictable ways. Particularly when dealing with texts as instantiated by paper technology, the reader needs to be aware of both the physical organization of the book and the logical structure of the work it contains. (Sperberg-McQueen and Bernard 1994)

The Elsevier DTD defines document structure in two ways. According to its creators, a full-length article has both an "article" structure, which is considered "coarse-grained," and a "text" structure which is considered "fine-grained" (Poppelier, van der Togt and Veldmeijer 1996, 38). The article structure is defined in the DTD as a sequence of elements, as indicated in Figure 1, based on a perceived model of scholarly discourse: a copyright statement, followed by headers and frontal matter, the body of the article, and then back matter consisting of references, further reading, an author's vita, and so on. However, within that structure, we have "text" structure, which is defined by parameter entities in the DTD, and provides a range of features that can occur anywhere within the carefully-defined hierarchy of the article. "Text," for instance, can be either a paragraph ("p") or a section (sec"). "%data" can consist of parsed-character data, or of text in different fonts, of footnotes, anchors and targets of hypertext links, tables or figures; thesecan appear in any sequence and any order, as indicated by the connectors in the entity declaration (see Figure 1).

Figure 1. Parameter Entity for Text Which Can Occur Within a Paragraph.


```
<!ENTITY % data    "(#PCDATA | %font-change; | %inline; | %display; |
                    fn | anchor | cross-ref | intra-ref | inter-ref)*">
```

What is notable about this "text" structure is that it confounds both hierarchy and sequence; the features of text can appear anywhere, in any order. Within the ordered hierarchy of the article is a less-ordered, diverse mish-mash of elements known as "text."

Levels of Discourse: Functions, Actions and Narrations

Having posited a view of narrative as a hierarchical paradigm with integrational and distributional planes of relationship, Barthes proceeds to establish a "provisional profile" of narrative structure which consists of three separate levels:

> *functions*: the smallest narrative units which serve a narrative function
>
> *actions*: the collection of functions into the operations of characters, or "actants"
>
> *narration*: the integration of actions into a narrative discourse

Barthes' ensuing discussion of these three levels reveals distinctive and suggestive features of narrative structure: features which, while unique to narrative, suggest similar features present in scholarly discourse.

Functions

In his isolation of the smallest units of narrative, Barthes classifies these units as distributional or integrational functions; nuclei and catalyzers work within a level of meaning, while indices and informants forge connections across levels. Barthes' taxonomy resists translation in this discussion; while Barthes maintains that "narra-

tive" can encompass forms from stained-glass windows to news reports, he is clearly limiting himself at this point to narrative in the traditional sense.[3] However, Barthes's discussion of the "functions" of narrative units employs two major concepts that are useful to us. First, he distinguishes between the units which are necessary to the narrative and those which are optional. Narrative, he argues, requires its *nuclei*: its moments in the narrative (generally paired), in which an alternative of direct consequence to the story is presented and then resolved. Narrative consists of sequences and sub-sequences of nuclei, with expansions provided both along the distributional level (in the case of catalyzers) or between levels (in the case of indices and informants): "Nuclei," he declares, "are at once necessary and sufficient. Once the framework they provide is given, the other units fill it out according to a mode of proliferation in principle infinite" (1966, 269). In a DTD, therefore, the distinction between mandatory and optional elements may well define the possibilities of a specific discourse.

Second, Barthes notes that nuclei can be important without being spectacular: "At first sight, such functions may appear extremely insignificant" (1966, 266): he cites the example of a telephone ringing and being answered. In narrative, then, structurally important units do not necessarily require overt prominence; they may not even require a specific position in an overall structure. They are often a function of content, and, in SGML terms, would appear in the document instance, without being anticipated by the preceding structure.

If we examine the Elsevier DTD with these two points in mind, some suggestive featurescome to light. First, the number of mandatory elements in a document instance based on this DTD is surprisingly low. In the root element, only the copyright statement is mandatory: everything else, including the front matter, body and back matter, are labelled as optional. This flexibility probably has less to do with any concept of necessary elements in scholarly discourse, and more to do with making one DTD fit several purposes; the attribute list for the "article" element indicates that the article DTD can be used for an abstract, addendum, announcement, calendar, conference, correspondence, editorial, erratum, literature alert

and patent, among many other options as well. The Elsevier DTD, then, sacrifices discursive coherence for flexibility in its structure; the mandatory or optional status of its elements is determined not by the requirements of a particular discourse, but by the need to make one structure fit different discourses.

Second, the presence of large, flexible parameter entities suggests that discursively-significant units can appear virtually anywhere in the document, and have not had their positions and relationships defined in the DTD. Among the many qualities of text provided in the parameter entities, one can recognize at least two different classes of elements which could indicate the presence of isolated or connected discursive units:

> *emphasis*: free-floating elements which enable the author to highlight text as bold, italic or underlined, or in some other fashion. The Elsevier DTD provides a notable example of this with its "catch-all" element, "Enunciation," which is used for articulating theorem, proof and definition statements (Poppelier 1996, 39).
>
> *links:* elements which enable the author to identify anchors and targets, and thereby provide connections to semantically- or discursively-significant units, either within or outside the document.

The Elsevier DTD allows for significant units to be tagged, identified, emphasized and linked wherever they appear in the article structure. Given the free-floating nature of these elements and the anti-hierarchical bias of hypertext,[4] one can surmise that much of scholarly discourse is constructed in ways that elude precise definition and placement in a DTD. The value of a DTD for a scholarly article, therefore, lies partly in its ability, not to define specific hierarchies, but to provide the freedom and flexibility needed for such hierarchies to emerge on their own.

Actions

Like the discussion of functions, Barthes's discussion of "actions" also needs to be considered skeptically, since by actions he is consid-

ering chiefly characters in a narrative. Barthes's analysis, furthermore, is partly motivated by a resistance to critical and readerly tendencies to read characters as psychological entities, who resemble real people and can be analyzed as such. We can analyze the structure of a DTD for scholarly articles without being drawn into that particular controversy; nonetheless, some of Barthes's conclusions have a relevance for our discussion. First, Barthes argues that a structural analysis of narrative sees characters as participants in actions rather than as beings: "Structural analysis . . . has so far striven . . . to define a character not as a 'being' but as 'participant' (1966, 277). This distinction has relevance to a profession which is currently renegotiating its paradigms of information objects and information things[5]: Barthes's treatment of character suggests that one can read an information object, such as a scholarly article, as a process rather than a thing, and that the multiple elements in a DTD are participants in an activity, rather than passive objects in a static structure. Second, Barthes argues that characters in a narrative need to be considered grammatically, rather thanpsychologically: that their significance arises not from their psychological essence, but from their relationship within the paradigmatic and syntagmatic structure of the narrative.

The Elsevier DTD, like all DTDs, defines its elements according to a rigorous grammar, which dictates when and how often an element may appear; if the DTD is to parse properly, this grammar, which is manifested in the hierarchical organization, the occurrence indicators and the frequency indicators, must be followed scrupulously.

Furthermore, a close examination of the elements in the front matter of a full-length article indicates that an Elsevier article, like all scholarly communication, is perceived largely as a communication act: the title of the article is followed by statements of authorship, collaboration and dedication, information about the reception, and acceptance and revision of the article. Whatever the status of the article may be as an information "object," the DTD structure clearly reflects a perception of the article as an act of scholarly communication, an act in which the various elements defined in the DTD are participating according to a rigid grammatical structure which is unconnected by the semantic content of any single document instance.

Narratives

The status of scholarly publishing as an act of communication is further enhanced when we consider Barthes's section on "Narration," in which the concept of communication and exchange is made explicit: "Just as there is within narrative a major function of exchange . . . so, homologically, narrative as object is the point of a communication: there is a donor of the narrative and a receiver of the narrative" (1966, 280-81). With this recognition of the document as the node of a communication act comes the recognition of several important concepts.

First, Barthes's insistence that communication depends on an exchange between donor and receiver ties into an observation made by the creators of the Elsevier Science DTD: namely,that some sorts of material are too easily misinterpreted to be presented according to SGML principles. Mathematical notation, of the sort that appears in scientific publishing, is incomprehensible without a shared knowledge base:

> Mathematical notation is designed to create the correct ideas in the mind of the reader. It is *deliberately* ambiguous and incomplete . . . the intrinsic information content of any mathematical formula is very low. A formula gets its meaning . . . only when used to communicate between two minds which share a large collection of concepts and assumptions, together with an agreed language for communicating the associated ideas. (Poppelier, van Herwijnen and Rowley 1992, 5)

This ambiguity means that there are multiple ways of presenting mathematical formulae in a DTD: so many, in fact, that the authors of the Elsevier DTD adopted a notation that emphasizes presentation, rather than logical structure.

Most important of all, however, is Barthes's discussion of narrative codes: "the set of operators which reintegrate functions and actions in the narrative communication" (1966 285). Our society, Barthes argues, repeatedly attempts to hide its narrative codes, "feigning to make [the narrative] the outcome of some natural circumstance"

(1966, 287). For Barthes, comprehension of a particular discourse rests upon the establishment of the proper codes in both the donor and the receiver; these codes establish the presuppositions necessary for the document to be "read" in the correct way. The back matter of the Elsevier DTD suggests an awareness that scholarly communication exists within a specific context which must be incarnated, to some extent, within the individual article (Figure 2):

Figure 2. The element for "back matter" in the Elsevier DTD. ©Elsevier Science Publishing

```
!ELEMENT bm    - O    (ack?, appn?, bibl*, further-reading*, glossary*, vt*)>
```

The provision of a background context is a fundamental part of the article's persuasiveness and its ultimate usefulness to the reader: the act of scholarly communication embodied in the article unfolds against a backdrop of acknowledgements, of appendices for further information and illustration, a collection of bibliographic references, and a description of the author's professional credentials. These elements are useful in their own right, providing amplification and directions for future research; they also serve the rhetorical purpose of persuading the reader through the structured display of a context of authority.

But there is more to narrative code than this. Barthes argues that some recognition of codes is taking place at the commencement of the reading process, and that this process of "loading" the codes is essentially a capacity to recognize and distinguish different genre of reading material, and the reading strategies required for each:

> however familiar, however casual may today be the act of opening a novel or a newspaper or of turning on the television, nothing can prevent that humble act from installing in us, all at once and in its entirety, the narrative code we are going to need. (1966, 287-88)

Barthes's statement is reinforced by the findings of HCI research, which has revealed the importance of interface metaphors as a means of enabling users to draw on prior knowledge to understand and

manipulate a system.[6] But such findings have discovered that the "codes" of recognition that are loaded by the user upon commencing the interaction with a text are features of layout as well as of structure, of formatting as well as of form. If this is the case, then many of the discursive functions which Barthes's discusses in "narration" occur not at the level of DTDdesign, nor at the occasion of a document instance. Rather, they occur in the process of styling the document for presentation: either through the creation of stylesheets in a program like Panorama, or in the writing of Perl scripts to translate an SGML document into HTML for mounting on the Web. When discussing the exchange of narrative "codes," to use Barthes's term, we can recognize the creation of discourse and discursive conventions beyond the explicit boundaries of SGML, within the concerns of text styling and interface design.

Conclusions

What, then, can we say about the Elsevier DTD, and SGML as a metalanguage, based on this discussion of Barthes? Primarily, we can recognize that the creation of discourse and discursive structure, on the level that Barthes is discussing, takes place partly, but not entirely within the structure of a Document Type Definition. While the DTD does indeed define a hierarchical structure of elements, with a rigorous grammar and a complex system of syntagmatic and paradigmatic relationships, it also proves its worth by providing flexibility, thereby enabling the discursive structure to manifest itself in the *parole* of the document instance, as well as in the *langue* of the DTD. As a result, a structuralist emphasis on the *langue* of a metalanguage like SGML can only be taken so far: sooner or later, we must adapt our conceptions of discourse along the lines of Bakhtin and Foucault, studying the individual utterances of text, rather than confining ourselves to the formal rules that govern their creation. Furthermore, the relationship between structure and interface needs to be explored: in particular, the tension between capturing the semantic units at the definition stage and rendering them visually to enable the correct communication of narrative codes.

End-Notes

1. Misiek and Oxford (1996) provide an account of the issues surrounding the creation of the SGML DTD for the FIS Electronic Journals Collection.

2. Rebecca Guenther (1996) provides a brief overview of the possibilities and limitations of SGML as both a supplement to and a replacement for US-MARC.

3. Although Barthes is confining himself to literary narrative, his choice of sample narrative–Ian Fleming's *Goldfinger*–is typically anti-canonical.

4. For a discussion of hypertext's ability to counteract traditional hierarchies of meaning, see George Landow's discussion of hypertext as "an infinitely decenterable and recenterable system" (1991, 17).

5. Linda Schamber's recent study distinguishes the electronic document from a printed document by its manipulable, transformable and transportable character, making it something "transient, fluid, constantly evolving" (1996, 670).

6. Microsoft's "desktop" metaphor is probably the most familiar to today's users. Other interface metaphors that have been tried include the "house" metaphor (Pejtersen and Nielsen 1991) and the electronic city metaphor (Toms and Kinnucan 1997).

References

Barate, Jean-Michel. 1994. "Roland Barthes." In *The Johns Hopkins Guide to Literary Theory and Criticism*. Baltimore: Johns Hopkins University Press 68-72.

Barthes, Roland. 1966. "Introduction to the structural analysis of narratives." In *A Barthes Reader*, ed. Susan Sontag. New York: Hill and Wang, 1982 251-295.

Beghtol, Clare. 1986. "Bibliographic classification theory and text linguistics: Aboutness analysis, intertextuality and the cognitive act of classifying documents." *Journal of Documentation* 42: 84-113.

Dillon, George L. 1994. "Discourse theory." In *The Johns Hopkins Guide to Literary Theory and Criticism*. Baltimore: Johns Hopkins University Press 210-212.

Elsevier Science Full-Length Article DTD. 1993-1997. Version 4.1.0. Elsevier Science. Internet WWW page, at URL: <http://www.elsevier.com/homepage/about/sgml/article/art410.dtd>

Guenther, Rebecca. 1996. "The challenges of electronic texts in the library: Bibliographic control and access." In *Scholarly Publishing: The Electronic Frontier*, Ed. Robin P. Peek and Gregory B. Newby. Cambridge: MIT Press 251-275.

Landow, George P. 1992. *Hypertext: The Convergence of Contemporary Critical Theory and Technology*. Baltimore: Johns Hopkins University Press.

Macleod, Ian A. 1990. "Storage and retrieval of structured documents." *Information Processing and Management* 26: 197-208.

Misiek, M., & Oxford, G. 1996. "The Electronic Resources Project of the Faculty of Information Studies." In *Proceedings of the 24th Annual Conference of the Canadian Association for Information Science. June 2-3, 1996*, 73-91.

Toronto, Ontario. Internet WWW page, at URL: <http://www.fis.utoronto.ca/ejournals/ erp-cais.html>

Pejtersen, A.M. and F. Nielsen. 1991. "Iconic interface for interactive fiction retrieval in libraries based on cognitive task analysis." In *Human Aspects in Computing Design and Use of Interactive Systems and Work with Terminals*, ed. H.J. Bullinger. Amsterdam: Elsevier 753-762.

Poppelier, N.A.F.M., E. van Herwijnen, and C.A. Rowley. 1992. *Standard DTDs and Scientific Publishing.* PDF file, at URL: <http://www.elsevier.com/homepage/about/sgml/docs/ epsig_news_article.pdf>

Poppelier, N.A.F.M., H. van der Togt, and F.K. Veldmeijer. 1996. *Document Type Definitions for Serial Publications.* Elsevier Science. PDF file, at URL: <http://www.elsevier.com/homepage/about/ sgml/docs/refman.pdf>

Rubinsky, Yuri. 1990. "Foreward." In *The SGML Handbook* by Charles Goldfarb. Oxford: Clarendon ix-xii.

Schamber, Linda. 1996. "What is a document? Rethinking the concept in uneasy times." *Journal of the American Society for Information Science* 47: 669-671.

Sperberg-McQueen, C.M. and Lou Bernard, Ed. 1994. "A gentle introduction to SGML." *Guidelines for Electronic Text Encoding and Interchange (TEI P3).* Text Encoding Initiative. Internet WWW page, at URL: <http://ota.ahds.ac.uk/ota/teip3sg/>

Toms, Elaine G. and Mark T. Kinnucan. 1996. "The effectiveness of the electronic city metaphor used to organize the menus of free-nets." *Journal of the American Society for Information Science* 47: 919-931.

van Herwijnen, Eric. 1994. *Practical SGML.* 2nd ed. Boston: Kluwer Academic Publishers.